Paper 140

Integrating SAS® Products Using Common Meta-Data

Catherine A. Phipps, SAS Institute Inc., Cary, NC

ABSTRACT

SAS® software products have always kept meta-data, or data about their data. But it was not until the introduction of a 'common meta-data model' that these products have actually been able to share their meta-data. SAS/EIS® software, SAS/MDDB™ software, SAS/Warehouse Administrator™ software and others have or are adding support for a common meta-data model. Some examples of the closer integration of these products through their common meta-data layer are discussed. For example, the ability to seamlessly create meta-data in one product and exploit it in another product is presented.

INTRODUCTION

Most SAS software products define and manipulate meta-data of some sort. In the past, this meta-data was the exclusive "property" of that application. However, the integrated meta-data strategy now being adopted at SAS Institute, using a common meta-data model, will enable these formerly disjoint applications to interoperate and add value to each other.

A META-DATA OVERVIEW

First of all, it is important to have a common understanding of what "meta-data" actually is. The most common definition is "data about data". An important example of this is a database table: What is its name? What type of information does it contain? How many columns does the table have? What are the names of the columns? the data types?

A data warehouse administration tool, such as the SAS/Warehouse Administrator software describes not only the physical structure of the data, but also the mappings, transformations and processes needed to turn the transactional data into the data warehouse.

This description of "physical meta-data" is an important aspect of meta-data in general, but it is by no means the only aspect. Meta-data also describes the structure of an organization and the business rules by which it runs. It can establish enterprise-wide standards for describing entities within the organization. That is, when an "employee" entity is described to the organization as part of its "logical" meta-data, each department that needs to process information about employees will be acting on a consistent description of what an "employee" means, even if the physical stores are different.

Application meta-data is yet another important kind of meta-data. This meta-data provides a consistent environment with which an application can process its data. The list of user ids authorized to create a certain report is an example of this application meta-data. The list of defined user attributes in a SAS/EIS software attribute dictionary is another example of application meta-data.

Although meta-data can describe many different things and can serve many purposes, there is one key element to recognize: Meta-data, if used effectively, can be the glue that binds disparate pieces of an organization together, binds physically separate databases together, and enables different applications to integrate with each other

THE COMMON META-DATA MODEL

The structure and organization of meta-data is called the "meta-data model". It consists of the entities being described, the attributes of those entities, and the ways in which those entities relate to each other. It is very much like a data model in that sense, except that the "data" in this case is meta-data.

All meta-data is part of a meta-data model of some kind. Not unlike urban sprawl, that "model" often grows and expands without apparent direction or coherence. What generally happens in the software industry in general is that each application or tool will create and manipulate its own meta-data without regard to what has already been done in another application. The result is a profusion of smaller meta-data models, each differing from each other in certain ways and none of them communicating and adding value to any others.

Yet, if these disparate meta-data models could be consolidated into a single "common" meta-data model, the benefits are immediate and far-reaching. Each application would be able to understand the meta-data produced by another application, so that effort spent defining meta-data in one application would enable it to be defined for other applications as well. The definition is then consistent across applications because it is a single definition. Application developers do not need to spend time defining a meta-data model specific to that application, so application development effort can be reduced or redirected to other important areas.

Unfortunately, achieving a common meta-data model is never as easy as the simple edict to "Make It So". In reality, defining a common model is a very difficult thing to do despite the best intentions of those involved. Existing applications already have a defined meta-data model that is especially customized to the needs of that application. Bringing together the commonality of multiple meta-data models always involves compromises of some sort. A "similar" meta-data model is by definition not the "same" meta-data model. It involves an incentive to change what was working well to embrace a new model that may be as yet unproven.

COMMON META-DATA MODEL "STANDARDS"

Yet, the advantages of a common meta-data model are so compelling that vendors throughout the software industry are coming together (albeit in different forums) to define an industry-wide common meta-data model.

Unfortunately, there is as yet no clear "winner" in the competition for a common model definition. One of the contenders is, of course, Microsoft with its OIM (Open Information Model), supported by its Repository product. Several vendors, including SAS Institute, have announced compatibility with that model in one form or another.

Another player in the common model game is the Meta Data Coalition with its MDIS (Meta Data Interchange Specification). SAS Institute has been, and continues to be, an active contributor to the development of MDIS. In the past several months, the focus of MDC has shifted a bit with the inclusion of Microsoft in the Coalition. There are on-going discussions about a merge of the Microsoft OIM with MDIS.

Yet another group of venders seeking to standardize meta-data management is OMG. The group has specified XMI (XML Meta-data Interchange), which unifies three important industry standards relating to meta-data and repositories: the Meta Object Facility (MOF) for distributed repositories, Unified Modeling Language (UML) to consistently describe meta-data models, and the XML protocol for web based meta-data management.

EXTENSIBLE MARKUP LANGUAGE (XML)

XML, or eXtensible Markup Language provides a way to represent the hierarchical nature of metadata by defining the document type definition tags (DTDs) to represent metadata objects. XML is a markup language that provides for definition of hierarchical definitions and object linking. It is well suited for the definition of metadata objects.

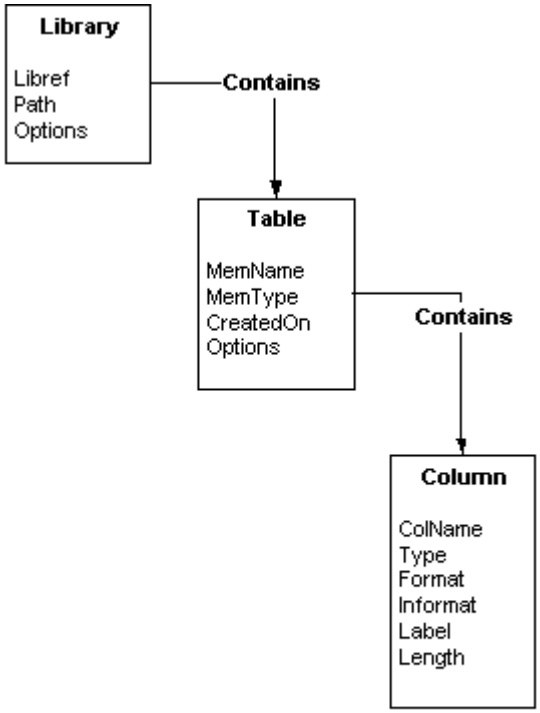
SAS software products will support interoperability to other repositories via an import/export facility, either through its support of MDIS or possibly, a more direct interchange.

TOWARDS A COMMON META-DATA STRATEGY

SAS Institute is actively pursuing an industry-wide common meta-data model. In the meantime, the Institute is proceeding with the development of a common meta-data strategy with its own products. In truth, SAS Institute products in Version 6 are a microcosm of the software industry in general, i.e., multiple products creating and using meta-data but unable to share the meta-data or their models. Integration occurs via an import/export facility, between specific products, e.g. SAS/Warehouse Administration software and SAS/EIS software.

In Version 7, the first steps towards truly integrated meta-data were taken with the introduction of CMR, the Common Meta-data Repository. CMR itself is not common meta-data, nor is it a common meta-data manager. CMR is a common meta-data facility. It provides a common set of meta-data services to SAS applications. Now SAS applications have the ability to store and retrieve their meta-data in a common format. The first SAS applications to use CMR in Version 7 are SAS/EIS software, SAS/MDDB software (HOLAP) and the SAS external file interface (EFI).

A common meta-data facility like CMR provides the building blocks needed for the truly integrated, shared meta-data. The other important piece is the development of a Common Meta-data Model (CMM) within the SAS System. This work is on-going and more SAS products will come on board with CMR, sharing meta-data via the CMM, in future releases of the SAS System. Here is a simplified diagram of the CMM for Version 7 of the SAS System.

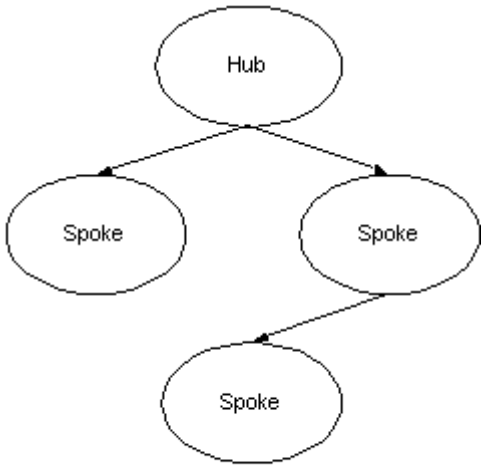


Version 7 Common Metadata Model
(simplified)

To understand how this could work in future releases, let us take the example of OLAP meta-data. Several SAS products, such as SAS/Warehouse Administrator software, SAS/EIS software, SAS/MDDB software and the MDDB procedure, all create or use OLAP-related meta-data. SAS/Warehouse Administrator software could define MDDBs (multidimensional databases). These definitions could be used by the SAS/MDDB Server, and read by SAS/EIS software. The SAS/MDDB Server could provide Tensor (OLAP for OLE DB) meta-data to be read and understood by the MDDB procedure. Each product could add its own value to the meta-data, resulting in a very rich model available for exploitation by multiple applications.

"HUB-AND-SPOKE"

It is important to note that having integrated, shared meta-data does not necessarily mean a single copy of the meta-data. SAS Institute supports a "hub-and-spoke": framework for its meta-data. That is, meta-data may be defined, created and maintained in one central location (the "hub"), then copied to other locations ("spokes") for exploitation. This offers the advantages of shared, centralized meta-data while at the same time providing the performance and availability of local meta-data access.



"Hub and Spoke" framework

Note that this diagram depicts that a single repository can act as both a Hub and a Spoke and thus a complex, distributed metadata environment can be created and supported. All of the spokes are active. Spokes can be created as department specific hubs, user specific hubs or application specific hubs depending on usage requirements. Spokes can be other SAS repositories or non-SAS metadata stores.

META-DATA APPLICATION PROGRAMMING INTERFACES

The meta-data contained in SAS repositories is, or will be, available outside of SAS software product interfaces via two APIs, or application programming interfaces.

The meta-data currently created and managed by the SAS/Warehouse Administrator software is available via the Meta-data API. This API allows you to write SCL programs that read, write, or update the SAS/Warehouse Administrator software meta-data without going through the SAS/Warehouse Administrator interface. This API will be enhanced to access meta-data in the Common Metadata Repository in a future release of the SAS System.

Another API currently under development opens up the meta-data on the SAS server to Enterprise Guide (MFC) and Java client applications. This API will be a counterpart to the existing SCL meta-data API, but will be using an XML-based protocol. We are looking at the XMI (XML Meta-data Interchange) format of the OMG group, possibly with extensions for SAS-specific meta-data requirements.

CONCLUSION

Meta-data is more than just “data about data”. It is the glue that, if used effectively, can bind together otherwise disparate applications. The keys to effective use of meta-data are 1) the definition of a common meta-data model, which will allow meta-data to be shared, with value added by each participating application2) interoperability between application meta-data stores, whether via a common meta-data facility such as CMR or via an import/export mechanism; and 3) a generalized meta-data API to open up the meta-data to both client and server-based applications. The integrated meta-data strategy of SAS Institute effectively addresses each of these key areas.

REFERENCES

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Catherine A. Phipps
SAS Institute Inc.
SAS Campus Drive, R4464
Cary, NC 27513
Phone: (919) 677-8000 x7757
Fax: (919) 677-4444
Email: sascap@wnt.sas.com